

**PERBANDINGAN METODE K-MEANS DAN FUZZY C-MEANS  
DALAM MENGELOMPOKKAN KABUPATEN/KOTA DI  
INDONESIA BERDASARKAN INDIKATOR INDEKS  
PEMBANGUNAN MANUSIA TAHUN 2021**

**SKRIPSI**



**Oleh  
BELIA MAILIEN  
NIM 18337033**

**PROGRAM STUDI SARJANA STATISTIKA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI PADANG  
2022**

**PERBANDINGAN METODE K-MEANS DAN FUZZY C-MEANS  
DALAM MENGELOMPOKKAN KABUPATEN/KOTA DI  
INDONESIA BERDASARKAN INDIKATOR INDEKS  
PEMBANGUNAN MANUSIA TAHUN 2021**

**SKRIPSI**

*Diajukan Sebagai Salah Satu Persyaratan Guna Memperoleh Gelar  
Sarjana Statistika*



Oleh  
**BELIA MAILIEN**  
**NIM 18337033**

**PROGRAM STUDI SARJANA STATISTIKA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI PADANG  
2022**

## PERSETUJUAN SKRIPSI

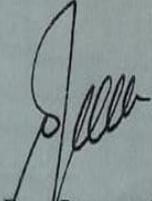
### PERBANDINGAN METODE K-MEANS DAN FUZZY C-MEANS DALAM MENGELOMPOKKAN KABUPATEN/KOTA DI INDONESIA BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN MANUSIA TAHUN 2021

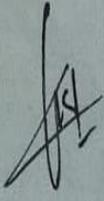
Nama : Belia Mailien  
NIM : 18337033  
Program Studi : SI Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 18 Agustus 2022

Mengetahui:  
Ketua Departemen Statistika

Disetujui Oleh:  
Pembimbing

  
Dr. Dony Permana, M.Si  
NIP. 197501272006041001

  
Admi Salma, S.Pd., M.Si  
NIDN. 0025129003

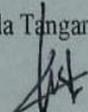
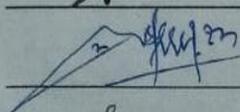
## PENGESAHAN LULUS UJIAN SKRIPSI

Nama : Belia Mailien  
NIM : 18337033  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

### PERBANDINGAN METODE K-MEANS DAN FUZZY C-MEANS DALAM MENGELOMPOKKAN KABUPATEN/KOTA DI INDONESIA BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN MANUSIA TAHUN 2021

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi  
Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Negeri Padang

Padang, 18 Agustus 2022

Tim Penguji	Nama	Tanda Tangan
Ketua	: Admi Salma, S.Pd., M.Si	
Anggota	: Dr. Syafriandi, M.Si	
Anggota	: Dina Fitria, M.Si	

## SURAT PERNYATAAN TIDAK PLAGIAT

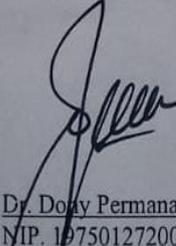
Saya yang bertandatangan di bawah ini:

Nama : Belia Mailien  
NIM : 18337033  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa, skripsi saya dengan judul **“Perbandingan Metode K-Means dan Fuzzy C-Means dalam Mengelompokkan Kabupaten/Kota di Indonesia Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2021”** adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan. Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan negara.

Demikian pernyataan ini saya buat dengan penuh kesadaran rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Diketahui oleh,  
Ketua Departemen Statistika,

  
Dr. Dody Permana, M.Si  
NIP. 197501272006041001

Saya yang menyatakan,



Belia Mailien  
NIM. 18337033

# PERBANDINGAN METODE K-MEANS DAN FUZZY C-MEANS DALAM MENGELOMPOKKAN KABUPATEN/KOTA DI INDONESIA BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN MANUSIA TAHUN 2021

**Belia Mailien**

## ABSTRAK

K-Means adalah salah satu analisis *cluster* non hirarki yang membagi objek ke dalam suatu kelompok berdasarkan jarak objek dengan pusat *cluster* terdekat. Fuzzy C-Means adalah analisis *cluster* non hirarki yang menggunakan model pengelompokan fuzzy dimana data dapat menjadi anggota suatu *cluster* yang terbentuk berdasarkan keanggotaan berbeda antara 0 sampai 1. Untuk mengetahui metode terbaik diantara kedua metode tersebut dilakukan pengelompokan Indeks Pembangunan Manusia (IPM). IPM merupakan indikator penting untuk mengukur keberhasilan upaya peningkatan kualitas hidup masyarakat. Peningkatan IPM di Indonesia tidak diiringi dengan pemerataan IPM di setiap kabupaten/kota di Indonesia. Agar terjadinya pemerataan IPM, pengelompokan dilakukan untuk mengetahui karakteristik dari tiap kabupaten/kota. Penelitian ini membahas tentang penggunaan algoritma K-means dan Fuzzy C-Means dalam mengelompokkan indikator IPM Tahun 2021 dengan jumlah *cluster* sebanyak 4.

Penelitian ini merupakan penelitian terapan yang dimulai dengan mempelajari teori - teori yang relevan dengan permasalahan. Data yang digunakan adalah data sekunder yang diperoleh dari *website* Badan Pusat Statistik (BPS) Republik Indonesia. Penelitian ini dimulai dengan mengelompokkan IPM indikator IPM Tahun 2021 dengan metode K-Means dan Fuzzy C-Means, kemudian dari hasil *cluster* yang diperoleh dilakukan uji validasi untuk mengetahui metode terbaik diantara kedua metode tersebut.

Hasil pengelompokan yang diperoleh adalah adanya indikator IPM yang rendah, sedang tinggi dan sangat tinggi. Nilai *C\_index* dan *S\_Dbw index* dari metode K-Means sebesar 0,108 dan 2,444. Sedangkan nilai *C\_index* dan *S\_Dbw index* dari metode Fuzzy C-Means sebesar 0,105 dan 2,312. Hal ini berarti hasil pengelompokan terbaik adalah menggunakan metode Fuzzy C-Means karena memiliki nilai *C\_index* dan *S\_Dbw index* yang lebih rendah.

Kata Kunci: Fuzzy C-Means, Indikator IPM, K-Means, Validasi *Cluster*

# COMPARISON OF K-MEANS AND FUZZY C-MEANS METHODS IN GROUPING DISTRICTS/CITIES IN INDONESIA BASED ON HUMAN DEVELOPMENT INDICATORS IN 2021

**Belia Mailien**

## **ABSTRACT**

K-Means is a non-hierarchical cluster analysis that classifies items depending on how far they are from the nearest cluster center. Data can be a member of a cluster that is produced based on various memberships between 0 and 1 using the fuzzy grouping model used by fuzzy C-Means, a non-hierarchical clustering method. The Human Development Index (IPM) is gathered to determine which technique is the best among the two. A key indicator of how well community initiatives to improve quality of life are doing is the HDI. There has been a growth in HDI in Indonesia, but it has not been distributed equally throughout the country's districts and cities. Grouping is done to identify the features of each district/city to ensure that HDI is distributed equally. The classification of the 2021 HDI indicators into 4 clusters using the K-means and Fuzzy C-means algorithms is covered in this work.

Studying theories that are pertinent to the issue is the first step in this research, which is an application of research. Secondary data were used, and they were found on the Central Statistics Agency of the Republic of Indonesia's website. This study starts by using the K-Means and Fuzzy C-Means methods to cluster the 2021 HDI indicators. From the cluster results, validation tests are run to compare the two methods and find which one is the most effective.

The presence of the low, medium, and extremely high HDI indicators is the outcome of the grouping achieved. The K-Means method's C index and S Dbw index values are 0.108 and 2.444, respectively. While the Fuzzy C-Means method's C index and S Dbw index values are 0.105 and 2.312, respectively. Due to its lower C\_index and S\_Dbw index values, the Fuzzy C-Means approach produces the best grouping results.

Keywords: *Cluster* Validity, Fuzzy C-Means, HDI indicator, K-Means

## KATA PENGANTAR



*Bismillahirrahmanirrahiim, Alhamdulillahirrabil'alamiin*, segala puji dan syukur penulis ucapkan kepada Allah SWT yang telah memberikan rahmat, kemudahan dan kesehatan kepada penulis sehingga dapat menimba ilmu kepada dosen-dosen yang memiliki keahlian di bidang Statistika. Dari ilmu yang telah diberikan, penulis dapat menyelesaikan skripsi dengan judul “Perbandingan Metode K-Means dan Fuzzy C-Means dalam Mengelompokkan Kabupaten/Kota di Indonesia Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2021”. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan Program Studi Statistika, Departemen Statistika, Fakultas Matematika dan ilmu Pengetahuan Alam, Universitas Negeri Padang. Dalam menyelesaikan skripsi ini penulis mendapatkan banyak bantuan dan dukungan dari berbagai pihak. Untuk itu penulis ucapkan terimakasih kepada:

1. Ibu Admi Salma, S.Pd., M.Si. dosen pembimbing skripsi yang telah memberikan arahan, bimbingan, dukungan dan motivasi dari awal sampai proses penyusunan skripsi.
2. Bapak Dr. Syafriandi, M.Si. dan Ibu Dina Fitria, M.Si. dosen penguji skripsi yang telah memberikan saran dan masukan positif dalam perbaikan skripsi penulis.

3. Bapak dan Ibu Dosen serta Staf Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang yang telah membantu penulis selama menimba ilmu di Program Studi Statistika.
4. Terkhusus orang tua penulis, Bapak Mulyadi dan Ibu Nofrida serta adik penulis yang telah mendukung secara materil dan non materil serta mendo'akan penulis tanpa henti dalam proses menimba ilmu pada Program Studi Statistika, Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang.
5. Rekan-rekan Program Studi S1 Statistika yang telah berjuang bersama selama perkuliahan.

Semoga skripsi ini memberikan manfaat untuk penulis sendiri, bermanfaat untuk semua pihak, dan bernilai ibadah di sisi Allah SWT. Skripsi ini tidak terlepas dari kesalahan dan kekeliruan, oleh sebab itu penulis mengharapkan saran dan kritik yang bersifat membangun. Akhir kata penulis ucapkan terima kasih.

Padang, 16 Agustus 2022

Penulis

## DAFTAR ISI

<b>ABSTRAK</b> .....	i
<b>ABSTRACT</b> .....	ii
<b>KATA PENGANTAR</b> .....	iii
<b>DAFTAR ISI</b> .....	v
<b>DAFTAR GAMBAR</b> .....	vi
<b>DAFTAR TABEL</b> .....	vii
<b>DAFTAR LAMPIRAN</b> .....	viii
<b>BAB I PENDAHULUAN</b> .....	1
A. Latar Belakang Masalah.....	1
B. Batasan Masalah Penelitian .....	6
C. Rumusan Masalah Penelitian .....	6
D. Tujuan Penelitian .....	6
E. Manfaat Penelitian .....	7
<b>BAB II KERANGKA TEORITIS</b> .....	8
A. Kajian Teori .....	8
B. Penelitian yang Relevan.....	20
<b>BAB III METODOLOGI PENELITIAN</b> .....	22
A. Jenis Penelitian.....	22
B. Jenis dan Sumber Data Penelitian .....	22
C. Variabel Penelitian.....	22
D. Teknik Analisis Data.....	23
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	28
A. Hasil .....	28
B. Pembahasan.....	38
<b>BAB V PENUTUP</b> .....	41
A. Kesimpulan .....	41
B. Saran .....	42
<b>DAFTAR PUSTAKA</b> .....	43
<b>LAMPIRAN</b> .....	46

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
1.Nilai IPM di Indonesia dari Tahun 2010-2021 .....	4
2.Diagram Alir Metode K-Means .....	23
3.Diagram Alir Metode Fuzzy C-Means.....	25
4.Boxplot Angka Harapan Hidup Menurut Kabupaten/Kota di Indonesia Tahun 2021 ...	28
5.Boxplot Harapan Lama Sekolah Menurut Kabupaten/Kota di Indonesia Tahun 2021..	29
6.Boxplot Rata-rata Lama Sekolah Menurut Kabupaten/Kota di Indonesia Tahun 2021.	30
7.Boxplot Pengeluaran per Kapita disesuaikan Menurut Kabupaten/Kota di Indonesia Tahun 2021.....	31
8.Sebaran Hasil <i>Cluster</i> Metode K-Means .....	34
9.Sebaran Hasil <i>Cluster</i> Metode Fuzzy C-Means.....	37

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
1. Kategori IPM .....	4
2. Variabel Penelitian.....	22
3. Struktur Data Penelitian.....	22
4. Hasil <i>Cluster</i> Metode K-Means .....	32
5. Karakteristik <i>Cluster</i> Berdasarkan Nilai Rata-Rata <i>Cluster</i> pada Metode K- Means..	33
6. Hasil <i>Cluster</i> Metode Fuzzy C-Means.....	35
7. Karakteristik <i>Cluster</i> Berdasarkan Nilai Rata-Rata Cluster pada Metode Fuzzy C- Means.....	36
8. Validasi Cluster K-Means dan Fuzzy C-Means.....	38

## DAFTAR LAMPIRAN

<b>Lampiran</b>	<b>Halaman</b>
1. Data Indikator IPM Tahun 2021 Menurut Kabupaten/Kota Tahun 2021 .....	46
2. Hasil Standardisasi Data .....	48
3. Hasil <i>Cluster</i> K-Means dan Fuzzy C-Means .....	50
4. <i>Syntax RStudio</i> .....	63

# BAB I

## PENDAHULUAN

### A. Latar Belakang Masalah

Analisis multivariat merupakan analisis yang dapat digunakan untuk melakukan analisis terhadap lebih dari dua variabel secara bersamaan. Analisis multivariat terbagi dua yaitu teknik dependen dan teknik interdependen. Teknik dependen dapat didefinisikan sebagai teknik yang memiliki dua jenis variabel yaitu variabel respon dan variabel penjelas. Beberapa analisis multivariat teknik dependen adalah analisis korelasi kanonik, MANOVA, regresi berganda, dan analisis konjoin. Sebaliknya, teknik interdependen adalah teknik di mana tidak ada variabel yang didefinisikan sebagai variabel respon atau variabel penjelas. Beberapa analisis multivariat teknik interdependen adalah analisis *cluster*, analisis faktor dan *Multidimensional scaling* (Hair *et al*, 2013:4-14).

Analisis *cluster* adalah proses pengelompokan sekumpulan objek ke dalam beberapa kelompok sehingga objek-objek dalam sebuah kelompok memiliki kemiripan yang tinggi, tetapi sangat berbeda dengan objek di kelompok lain (Han *et al*, 2012:443). Algoritma pengelompokan lainnya adalah klasifikasi. Klasifikasi bertujuan untuk menetapkan pengamatan baru ke salah satu kelompok, sedangkan analisis *cluster* untuk mengelompokan objek atas dasar kesamaan atau jarak (*dissimilarities*) agar diketahui kesamaan dari objek tersebut (Jonhson dan Winchern, 2002: 688).

Analisis *cluster* secara umum terbagi menjadi dua yaitu metode hirarki dan non hirarki. Metode hirarki digunakan untuk mengelompokkan objek secara terstruktur berdasarkan kemiripan sifatnya dan jumlah *cluster* yang diinginkan

belum diketahui banyaknya. Metode hirarki terbagi 2 yaitu *agglomerative* (penggabungan) dan *divisive* (pemecahan). Bagian *agglomerative* terdiri dari metode *Single Linkage*, *Complete Linkage*, *Average Linkage* dan *Ward* (Mattjik dan Sumertajaya, 2011:199). Menurut Jonhson dan Winchern (2002: 694) metode non hirarki mengelompokkan objek kedalam suatu kelompok dimana jumlah kelompok ditentukan terlebih dahulu. Metode non hirarki lebih baik dalam menangani kumpulan data yang besar dibanding metode hirarki. Beberapa metode analisis *cluster* non hirarki seperti metode K-Means, Fuzzy C-Means, K-Medoids atau PAM (*Partitioning Around Medoids*), dan CLARA (*Clustering Large Applications*).

K-Means adalah salah satu analisis *cluster* non hirarki yang membagi objek kedalam kelompok *cluster* berdasarkan jarak objek dengan pusat *cluster* terdekat (Jonhson dan Winchern, 2002: 694). K-Means sangat sederhana, efisien, cepat serta dapat menangani dengan sangat efisien kumpulan data yang besar (Wu, 2012:8). Fuzzy C-Means adalah analisis *cluster* yang menggunakan model pengelompokkan *fuzzy* dimana data dapat menjadi anggota suatu *cluster* yang terbentuk berdasarkan derajat keanggotaan berbeda antara 0 sampai 1. Fuzzy C-Means memberikan hasil yang baik jika terjadi tumpang tindih pada *cluster* (Cebeci dan Yildiz, 2015). Metode K-Means dan Fuzzy C-Means cocok digunakan untuk data dengan tipe variabel kontinu (Lathifaturrahmah, 2014).

Menurut Ross (2010: 340), ada istilah *hard partition* dan *soft partition*. K-Means dikenal dengan *hard partition* karena setiap amatan tepat berada dalam satu *cluster*, sedangkan Fuzzy C-Means dikenal dengan *soft partition* karena setiap amatan ditugaskan ke satu *cluster* tetapi diizinkan untuk berada dalam keanggotaan

parsial (King, 2015:137). Perbandingan metode K-Means dan Fuzzy C-Means telah dilakukan oleh beberapa peneliti sebelumnya. Sivarathri dan Govardhan (2014) pada pengelompokan data penyakit diabetes mengatakan bahwa akurasi dan kualitas *cluster* yang dibuat oleh Fuzzy C-Means lebih baik dari pada K-Means. Panda *et al* (2012) pada pengelompokan data Iris mengatakan bahwa dari segi sensitivitas, spesifisitas, presisi, akurasi, run time, jarak rata-rata intra *cluster* dan jarak antar *cluster* K-Means mengungguli Fuzzy C-Means.

Berdasarkan pemaparan tersebut, maka penulis tertarik untuk mengkaji lebih lanjut mengenai perbandingan metode K-Means dan Fuzzy C-Means yang memberikan hasil *cluster* terbaik, yaitu hasil *cluster* yang mempunyai variansi di dalam *cluster* yang lebih homogen dan variansi antar *cluster* yang lebih heterogen dengan permasalahan Indeks Pembangunan Manusia (IPM) di Indonesia Tahun 2021. IPM merupakan indikator penting untuk mengukur keberhasilan upaya peningkatan kualitas hidup masyarakat. IPM juga menjadi tolak ukur kemajuan suatu daerah, karena dengan IPM yang tinggi maka suatu daerah akan dianggap berhasil dalam program pembangunannya. Untuk membangun IPM yang berkualitas, pemerintah harus memiliki program dan fasilitas yang memadai untuk meningkatkan IPM serta dukungan masyarakat agar program tersebut dapat berjalan.

Menurut *United Nations Development Programme* (UNDP), IPM terdiri dari tiga dimensi dasar yaitu umur panjang dan hidup sehat, pengetahuan, dan standar kehidupan yang layak. Dimensi umur panjang dan hidup sehat digunakan indikator Angka Harapan Hidup (AHH), dimensi pengetahuan digunakan gabungan indikator Harapan Lama Sekolah (HLS) dan Rata-Rata Lama Sekolah (RLS), dan untuk

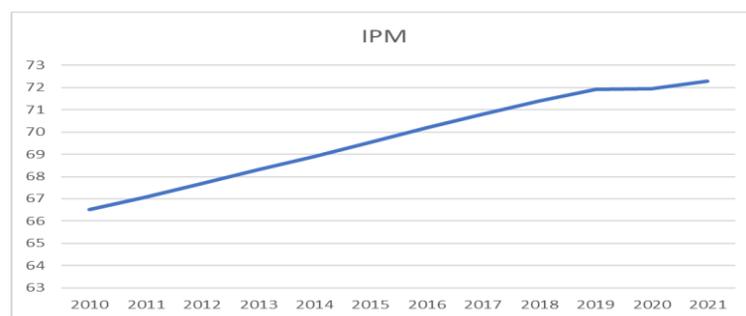
mengukur dimensi standar hidup layak digunakan indikator kemampuan daya beli masyarakat yang dilihat dari rata-rata besarnya Pengeluaran per Kapita (PPP) yang disesuaikan.

Sejak tahun 2014 angka IPM di Indonesia disajikan setiap tahun pada tingkat nasional, provinsi, dan kabupaten/kota. Pencapaian pembangunan manusia di suatu daerah pada waktu tertentu dapat dikelompokkan menjadi empat kelompok sesuai dengan status pencapaiannya. Adapun pengelompokan IPM dapat dilihat pada Tabel 1.

Tabel 1. Kategori IPM

Kategori	Nilai IPM
Rendah	$< 60$
Sedang	$60 \leq \text{IPM} < 70$
Tinggi	$70 \leq \text{IPM} < 80$
Sangat Tinggi	$\text{IPM} \geq 80$

Data yang tercatat di BPS menunjukkan IPM di Indonesia secara garis besar mengalami kenaikan setiap tahunnya. Kenaikan tersebut dapat dilihat pada Gambar 1.



Gambar 1. Nilai IPM di Indonesia dari Tahun 2010-2021

Berdasarkan Gambar 1, kenaikan IPM setiap tahun di Indonesia berarti Indonesia terus mengalami kemajuan dalam pembangunan. Namun jika diperhatikan penyebarannya pada tiap kabupaten/kota di Indonesia, tidak seluruhnya memiliki nilai IPM yang tinggi. Keberagaman nilai IPM menunjukkan

kesenjangan yang mencolok antara wilayah bagian timur dan wilayah bagian barat Indonesia (BPS, 2021).

Menurut Zusanti *et al* (2018), rendah atau tingginya IPM mempengaruhi tingkat produktivitas penduduk. Semakin rendah IPM, semakin rendah tingkat produktivitas penduduk, dan semakin rendah produktivitas, semakin rendah pendapatan. Sebaliknya semakin tinggi IPM maka semakin tinggi pula tingkat produktivitas penduduk dan semakin tinggi pula tingkat pendapatannya. Nilai IPM yang tinggi jika tidak diikuti dengan pemerataan pendapatan akan mengurangi kesejahteraan seluruh masyarakat yang menyebabkan ketimpangan wilayah.

Keberagaman nilai IPM diakibatkan karena kabupaten/kota memiliki keadaan yang berbeda pada setiap indikator penentu nilai IPM. Perlu diperhatikan indikator penentu nilai IPM untuk mengetahui indikator yang menjadi permasalahan ataupun keunggulan suatu kabupaten/kota agar diperoleh karakteristik dari masing-masing kabupaten/kota yang mempengaruhi nilai IPM.

Untuk mempermudah pemerintah dalam membuat kebijakan dan perencanaan dalam mengatasi tidak meratanya IPM di Indonesia, perlu adanya pengelompokan kabupaten/kota di Indonesia berdasarkan indikator IPM. Berdasarkan hasil pengelompokan yang diperoleh, dilakukan penjabaran mengenai karakteristik masing-masing *cluster* sebagai identifikasi terhadap karakteristik kabupaten/kota berdasarkan indikator penentu nilai IPM Tahun 2021.

Berdasarkan pemaparan yang telah dibahas, maka penulis melakukan penelitian yang berjudul **"Perbandingan Metode K-Means dan Fuzzy C-Means dalam Mengelompokkan Kabupaten/Kota di Indonesia Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2021"**.

## **B. Batasan Masalah Penelitian**

Untuk menghindari pembahasan yang terlalu melebar, perlu adanya batasan pada penelitian ini. Adapun batasan pada penelitian ini adalah menggunakan data indikator IPM Tahun 2021 yang meliputi Angka Harapan Hidup (AHH) saat lahir, Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Pengeluaran per Kapita yang disesuaikan. Metode yang akan digunakan untuk mengelompokkan kabupaten/kota di Indonesia berdasarkan indikator IPM menggunakan analisis K-Means dan Fuzzy C-Means.

## **C. Rumusan Masalah Penelitian**

Berdasarkan uraian latar belakang, adapun rumusan masalah dari penelitian ini adalah sebagai berikut.

1. Bagaimana hasil pengelompokkan kabupaten/kota di Indonesia berdasarkan indikator IPM Tahun 2021 menggunakan metode K-Means dan Fuzzy C-Means?
2. Metode *cluster* mana yang lebih baik antara K-Means dan Fuzzy C-Means dalam mengelompokkan kabupaten/kota di Indonesia berdasarkan indikator IPM Tahun 2021?

## **D. Tujuan Penelitian**

Tujuan dari penelitian ini adalah untuk menjawab rumusan masalah dalam penelitian ini yaitu:

1. Untuk mengetahui hasil pengelompokkan kabupaten/kota di Indonesia berdasarkan indikator IPM Tahun 2021 menggunakan metode K-Means dan Fuzzy C-Means.

2. Untuk mengetahui metode terbaik antara K-Means dan Fuzzy C-Means dalam mengelompokkan kabupaten/kota di Indonesia berdasarkan indikator IPM Tahun 2021.

#### **E. Manfaat Penelitian**

Manfaat dari penelitian yang ingin dicapai adalah sebagai berikut.

1. Bagi peneliti, dapat menambah pengetahuan dan penguasaan materi dalam analisis K-Means dan Fuzzy C-Means serta sebagai pengalaman dalam melakukan penelitian.
2. Bagi pembuat kebijakan, diharapkan penelitian ini dapat dijadikan sebagai salah satu acuan untuk membuat keputusan dan perencanaan agar terjadinya pemerataan IPM di Indonesia.
3. Bagi peneliti selanjutnya, sebagai bahan rujukan dan perbandingan untuk melakukan penelitian lanjutan.

## BAB II

### KERANGKA TEORITIS

#### A. Kajian Teori

##### 1. Analisis Multivariat

Menurut Hair *et al* (2013:4-14) analisis multivariat merupakan analisis yang dapat digunakan untuk melakukan analisis terhadap lebih dari dua variabel secara bersamaan. Analisis multivariat terbagi dua yaitu teknik dependen dan teknik interdependen. Teknik dependen dapat didefinisikan sebagai teknik yang memiliki dua jenis variabel yaitu variabel respon dan variabel penjelas. Sebaliknya, teknik interdependen adalah teknik di mana tidak ada variabel yang didefinisikan sebagai variabel respon atau variabel penjelas.

Data pada analisis multivariat dengan  $p$  variabel dan amatan  $n$  dapat ditulis dalam bentuk matriks sebagai berikut.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

$\mathbf{X}$  berisi data yang terdiri dari semua pengamatan pada semua variabel (Jonhson dan Winchern, 2002:5-6).

##### 2. Analisis *Cluster*

###### a. Defenisi Analisis *Cluster*

Analisis *cluster* berkaitan dengan eksplorasi sekumpulan data untuk menilai apakah data dapat diringkas secara bermakna kedalam kelompok objek atau individu yang mirip satu sama lain dan apakah

sekumpulan data berbeda dari individu dalam kelompok lain (Everitt *et al*, 2011:13). Analisis *cluster* biasanya melibatkan setidaknya tiga langkah. Yang pertama adalah pengukuran kemiripan atau tidak kemiripan antar data. Langkah kedua adalah proses pengelompokan dimana amatan dipartisi menjadi beberapa *cluster*. Langkah terakhir adalah menginterpretasi hasil *cluster* (Hair *et al*, 2013:18). Menurut Han *et al* (2012:448-449), *cluster* yang baik adalah objek dalam suatu *cluster* terkait satu sama lain sedangkan objek antar *cluster* berbeda satu sama lain.

Untuk mengukur kemiripan objek atau individu dalam suatu kelompok digunakan beberapa ukuran jarak (Mattjik dan Sumertajaya, 2011:196). Menurut Everitt *et al* (2011: 49), ukuran jarak yang paling umum digunakan adalah jarak Euclidean dengan rumus sebagai berikut.

$$d_{ik} = \left[ \sum_{j=1}^p (x_{kj} - x_{ij})^2 \right]^{1/2} \quad (1)$$

Keterangan:

$d_{ik}$  = Jarak Euclid objek ke-i dan objek ke-k

$x_{kj}$  = Objek ke-k pada variabel ke-j

$x_{ij}$  = Objek ke-i pada variabel ke-j

$p$  = banyak variabel yang digunakan

Terdapat beberapa metode yang bisa digunakan dalam analisis *cluster*, seperti metode hirarki. Metode hirarki digunakan untuk mengelompokkan objek secara terstruktur berdasarkan kemiripan sifatnya dan jumlah *cluster* yang diinginkan belum diketahui banyaknya (Mattjik dan Sumertajaya, 2011:199). Hasil

pengelompokkan dalam metode hirarki ditampilkan dalam bentuk dendogram (Kassambara, 2017: 65). Metode analisis *cluster* yang lainnya yaitu metode non hirarki. Menurut Jonhson dan Winchern (2002:694) metode non hirarki mengelompokkan objek kedalam suatu kelompok dimana jumlah kelompok ditentukan terlebih dahulu. Karena matriks jarak (kesamaan) tidak harus ditentukan dan data awal tidak harus disimpan selama komputer dijalankan, metode non hirarki dapat diterapkan untuk kumpulan data yang jauh lebih besar daripada metode hirarki.

b. Analisis *Cluster* Non Hirarki

Beberapa metode yang sering digunakan pada metode non hirarki yang cocok untuk data dengan variabel kontinu adalah metode K-Means dan Fuzzy C-Means.

1) K-Means

K-Means adalah salah satu analisis *cluster* non hirarki yang membagi objek kedalam kelompok *cluster* berdasarkan jarak objek dengan pusat *cluster* terdekat (Jonhson dan Winchern, 2002:694). Menurut Singh *et al* (2013), berikut adalah langkah-langkah pada algoritma K-Means.

- a) Menentukan terlebih dahulu banyaknya *cluster* ( $g$ )
- b) Menentukan titik pusat *cluster* ( $v_{lj}$ ) yang dipilih secara acak.
- c) Menghitung jarak objek ke titik pusat ( $d_{ik}$ )

- d) Mengelompokkan objek berdasarkan jarak terdekat objek ke titik pusat *cluster*.
- e) Pusat *cluster* pada iterasi selanjutnya dapat dihitung menggunakan persamaan berikut.

$$v_{lj} = \left(\frac{1}{n_l}\right) \sum_{k=1}^{n_l} x_{kj} \quad (2)$$

Keterangan:

$v_{lj}$  = pusat *cluster* ke- $l$  pada variabel ke- $j$

$x_{kj}$  = objek ke- $k$  pada variabel ke- $j$

$n_l$  = banyak objek pada *cluster* ke- $l$

Proses iterasi dihentikan ketika sudah tidak terjadi lagi perpindahan objek antar *cluster*.

## 2) Fuzzy C-Means

Menurut Everitt *et al* (2011:242), dalam *clustering* fuzzy setiap objek memiliki derajat keanggotaan yang menunjukkan kekuatan keanggotaan di semua atau beberapa *cluster*. Pengelompokan fuzzy memiliki keunggulan utama yaitu keanggotaan untuk setiap objek yang diberikan menunjukkan apakah ada *cluster* terbaik kedua yang hampir sebagus *cluster* terbaik pertama, sebuah fenomena yang sering disembunyikan saat menggunakan teknik *clustering* lainnya. Derajat keanggotaan dapat diskalakan untuk berada di antara nol dan satu, dan kemudian dapat diartikan sebagai probabilitas. Selain diartikan

sebagai probabilitas, dalam Fuzzy C-Means, derajat keanggotaan menggambarkan seberapa dekat objek dengan titik pusat.

Derajat keanggotaan suatu objek ke- $k$  pada *cluster* ke- $l$  adalah sebagai berikut.

$$\mu_{lk} = \mu_{A_l}(x_k) \in [0,1]$$

dengan  $\mu_{lk}$  merupakan derajat keanggotaan objek ke- $k$  pada *cluster* ke- $l$  dan  $A_l$  adalah anggota himpunan *cluster* ke- $l$ . Berikut adalah langkah-langkah pada metode Fuzzy C-Means.

- a) Menentukan banyaknya *cluster* ( $g$ ), parameter pembobot keanggotaan ( $m'$ ) dan tingkat akurasi  $\varepsilon_L$  dan banyaknya iterasi ( $r$ ). Nilai  $m'$  memiliki rentang  $m' \in [1, \infty]$ . Jika nilai  $m' = 1$  maka algoritma Fuzzy C-Means mendekati algoritma K-Means.  $m'$  mengontrol tingkat pembagian keanggotaan *cluster*. Nilai  $m'$  disarankan bernilai 2.
- b) Menentukan matriks partisi awal  $\mathbf{U}^{(0)}$ . Matriks  $\mathbf{U}^{(0)}$  terdiri dari nilai derajat keanggotaan yang dibangkitkan secara acak.

$$\mathbf{U}_{g \times n}^r = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1n} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{g1} & \mu_{g2} & \dots & \mu_{gn} \end{bmatrix} \quad (3)$$

Keterangan:

$\mathbf{U}_{g \times n}^r$  = matriks partisi berukuran  $g \times n$  pada iterasi ke- $r$

$g$  = ukuran *cluster* yang digunakan

$n$  = banyaknya objek

$\mu_{lk}$  = merupakan derajat keanggotaan objek ke- $k$  pada  
*cluster* ke- $l$

- c) Menghitung pusat *cluster* ke- $l$  pada variabel ke- $j$  dapat dihitung menggunakan persamaan berikut.

$$v_{lj}^{(r)} = \frac{\sum_{k=1}^n (\mu_{lk}^{(r)})^{m'} \cdot x_{kj}}{\sum_{k=1}^n (\mu_{lk}^{(r)})^{m'}} \quad (4)$$

Keterangan:

$v_{lj}$  = pusat *cluster* ke- $l$  pada variabel ke- $j$

$\mu_{lk}$  = derajat keanggotaan objek ke- $k$  pada *cluster* ke- $l$

$r$  = iterasi

$m'$  = parameter pembobot keanggotaan

- d) Menghitung jarak objek ke pusat *cluster*. Setelah itu melakukan pembaruan elemen matriks partisi  $\mathbf{U}^{(r)}$  dengan elemen berikut.

$$\mu_{lk}^{(r+1)} = \left[ \sum_{i=1}^g \left( \frac{d_{lk}^{(r)}}{d_{ik}^{(r)}} \right)^{\frac{2}{m'-1}} \right]^{-1} \quad (5)$$

$d_{lk}^{(r)}$  merupakan Jarak objek ke- $k$  pada pusat *cluster* ke- $l$ .

- e) Konvergensi hasil pengelompokkan dapat ditentukan menggunakan kriteria berikut.

$$\max_{l,k} \left| \mu_{lk}^{(r+1)} - \mu_{lk}^{(r)} \right| \leq \varepsilon_L \quad (6)$$

Dengan  $\varepsilon_L$  merupakan tingkat akurasi yang ditentukan.

Apabila kriteria konvergensi telah tercapai, maka iterasi dihentikan. Namun, jika belum tercapai maka hitung

kembali pusat *cluster* sampai kriteria konvergensi tercapai (Ross, 2010:349-352).

c. Standardisasi Data

Standardisasi dilakukan apabila variabel memiliki perbedaan ukuran satuan (Alwi dan Hasrul, 2018). Menurut Han *et al* (2012:113-114), untuk algoritma berbasis jarak seperti analisis *cluster* standardisasi dilakukan agar nilai variabel yang besar tidak mempengaruhi nilai variabel yang kecil begitupun sebaliknya. Standardisasi dapat dilakukan dengan menggunakan persamaan berikut.

$$z_{kj} = \frac{x_{kj} - \bar{x}_j}{s_j} \quad (7)$$

Keterangan:

$z_{kj}$  : *z score* objek ke- $k$  variabel ke- $j$

$x_{kj}$  : objek ke- $k$  variabel ke- $j$

$\bar{x}_j$  : rata-rata variabel ke- $j$

$s_j$  : simpangan baku variabel ke- $j$

d. Interpretasi Analisis *Cluster*

Interpretasi suatu *cluster* merupakan penggambaran karakteristik masing-masing *cluster* yang menjelaskan perbedaan dari setiap *cluster* secara relevan. Ukuran yang bisa digunakan untuk proses interpretasi ini adalah menghitung nilai rata-rata setiap objek pada masing-masing *cluster* (Hair *et al*, 450: 2013). Menurut Talakua *et al* (2017), jika dilakukan standardisasi pada data maka cara interpretasi hasil *cluster*

juga dilihat dari nilai rata-rata *cluster* pada tiap variabel dimana tanda negatif pada nilai rata-rata *cluster* menandakan bahwa nilai rata-rata *cluster* pada variabel tersebut berada dibawah nilai rata-rata variabel pada data keseluruhan sebelum dipisah kedalam bentuk *cluster*. Sedangkan tanda positif menandakan bahwa nilai rata-rata *cluster* pada variabel tersebut berada diatas nilai rata-rata variabel pada data keseluruhan sebelum dipisah kedalam bentuk *cluster*.

e. Validasi Metode Analisis *Cluster*

Ukuran validasi sangat dibutuhkan untuk mengetahui seberapa baik hasil pengelompokan (Wu, 2012:5). Ukuran validasi yang bisa digunakan adalah *S\_Dbw index* dan *C\_index*.

1) *S\_Dbw index*

Menurut Legany *et al* (2006), *S\_Dbw index* menghasilkan keputusan yang tepat dalam mengidentifikasi hasil *cluster*. *S\_Dbw index* mengukur varians intra *cluster* dan kepadatan antar *cluster*. *S\_Dbw* dirumuskan sebagai berikut.

$$S_{Dbw} = Scatt + Dens_{bw} \quad (8)$$

dengan

$$Scatt = \frac{1}{g} \sum_{i=1}^g \frac{|\sigma(v_i)|}{|\sigma(x_j)|}$$

dan

$$Dens_{bw} = \frac{1}{g(g-1)} \sum_{i=1}^g \frac{density(u_{lm})}{\max\{density(v_l), density(v_m)\}}$$

Keterangan:

$S\_Dbw$  :  $S\_Dbw$  index

$Scatt$  : varians intra *cluster*

$Dens\_bw$  : kepadatan antar *cluster*

$g$  : ukuran *cluster* yang digunakan

$\sigma(v_l)$  : varians pusat *cluster* ke- $l$

$\sigma(x_j)$  : varians variabel ke- $j$

$u_{lm}$  : titik tengah *cluster* ke- $l$  dan *cluster* ke- $m$ .

Kriteria pengelompokan baik menggunakan  $S\_Dbw$  index adalah nilai yang lebih rendah karena menunjukkan keragaman dalam suatu *cluster* yang rendah dan kepadatan antar *cluster* yang rendah (Legany *et al*, 2006).

## 2) $C\_index$

Menurut Charrad *et al* (2014),  $C\_index$  mengukur selisih jarak maksimum dan minimum dalam *cluster* dan antar *cluster*.  $C\_index$  dirumuskan sebagai berikut.

$$C\_index = \frac{S_w - S_{min}}{S_{max} - S_{min}} \quad (9)$$

Keterangan:

$S_w$  = jumlah total jarak dalam *cluster*

$S_{min}$  = jumlah jarak minimum dalam *cluster* dan antar *cluster*

$S_{max}$  = jumlah jarak maksimum dalam *cluster* dan antar *cluster*

Kriteria pengelompokan baik menggunakan  $C\_index$  adalah nilai yang lebih rendah karena menunjukkan jarak dalam

*cluster* yang rendah dan jarak antar *cluster* yang berjauhan (Charrad *et al*, 2014).

f. R Packages Analisis K-Means dan Fuzzy C-Means

*Packages* yang digunakan untuk metode K-Means adalah *fpc* yang dibuat oleh Henning dan Imports (2015). *Packages* yang digunakan untuk metode Fuzzy C-Means adalah *ppclust* yang dibuat oleh Cebecci *et al* (2019). Untuk validasi *cluster* digunakan *packages ClusterCrit* yang dibuat oleh Desgraupes (2013).

g. Indeks Pembangunan Manusia (IPM)

Gagasan mengenai pembangunan manusia pertama kali diperkenalkan oleh *United Nations Development Programme* (UNDP) pada Tahun 1990 melalui laporan *Human Development Report* (HDR). Gagasan tersebut lebih dari sekedar menempatkan manusia sebagai input dari pembangunan tetapi menjadikan manusia sebagai tujuan akhir dari pembangunan yaitu menciptakan lingkungan yang memungkinkan bagi masyarakat untuk produktif sehingga dapat memiliki umur panjang dan hidup sehat, menguasai pengetahuan, dan memenuhi standar hidup yang layak. HDR memberikan rekomendasi pengukuran pembangunan manusia yang disebut sebagai Indeks Pembangunan Manusia (IPM).

Menurut BPS (2021), IPM memiliki manfaat yaitu menjadi salah satu indikator penting dalam melihat sisi lain dari pembangunan. Setiap indikator komponen penghitungan IPM dapat dimanfaatkan untuk mengukur keberhasilan pembangunan kualitas hidup manusia. Manfaat

lain dari IPM adalah sebagai salah satu indikator target pembangunan dan salah satu alokator dalam penentuan Dana Alokasi Umum (DAU). Selain itu, IPM juga digunakan sebagai salah satu indikator pengukuran kinerja utama Dana Insentif Daerah (DID) dalam mendorong peningkatan kesejahteraan masyarakat sekaligus sebagai salah satu indikator dalam pengalokasian DID.

Pada Tahun 2010, UNDP melakukan penyempurnaan terhadap perhitungan IPM yang dikenal dengan perhitungan IPM metode terbaru. Tahun 2014 Indonesia secara resmi mengaplikasikan metode baru perhitungan IPM menggunakan indikator sebagai berikut.

- 1) Angka Harapan Hidup (AHH) saat lahir merupakan rata-rata perkiraan lamanya waktu (dalam tahun) yang dapat dijalani oleh seseorang selama hidupnya.
- 2) Harapan Lama Sekolah (HLS) menggambarkan lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak dimasa yang akan datang.
- 3) Rata-rata Lama Sekolah (RLS) menggambarkan jumlah tahun yang digunakan oleh penduduk dalam menjalani pendidikan formal.
- 4) Pengeluaran per kapita (PPP) disesuaikan merupakan indikator pada dimensi standar hidup yang layak yang mampu mencerminkan kemampuan daya beli masyarakat terkait membelanjakan uangnya dalam bentuk barang maupun jasa (BPS, 2021).

IPM dihitung sebagai rata-rata geometrik dari indeks umur panjang dan hidup sehat, pengetahuan dan standar hidup layak. Indeks kesehatan terdiri dari indikator AHH, indeks pengetahuan terdiri dari RLS dan HLS, indeks standar hidup layak/pengeluaran terdiri dari indikator PPP yang disesuaikan. Pada setiap dimensi tersebut merupakan nilai indikator yang distandardisasi dengan nilai minimum dan maksimum. Rumus yang digunakan untuk menghitung indeks dari dimensinya tersebut adalah sebagai berikut.

1) Indeks Kesehatan

$$I_{kesehatan} = \frac{AHH - AHH_{min}}{AHH_{maks} - AHH_{min}}$$

Keterangan:

$I_{kesehatan}$  : Indeks kesehatan

$AHH$  : Angka harapan hidup

2) Indeks Pendidikan

$$I_{RLS} = \frac{RLS - RLS_{min}}{RLS_{maks} - RLS_{min}}$$

$$I_{HLS} = \frac{HLS - HLS_{min}}{HLS_{maks} - HLS_{min}}$$

$$I_{pendidikan} = \frac{I_{RLS} - I_{HLS}}{2}$$

Keterangan:

$I_{RLS}$  : Indikator rata-rata lama sekolah

$I_{HLS}$  : Indikator harapan lama sekolah

$I_{pendidikan}$  : Indeks pendidikan

## 3) Indeks Pengeluaran

$$I_{\text{pengeluaran}} = \frac{\ln(\text{pengeluaran}) - \ln(\text{pengeluaran}_{\min})}{\ln(\text{pengeluaran}_{\max}) - \ln(\text{pengeluaran}_{\min})}$$

Keterangan:

Pengeluaran : pengeluaran per kapita yang disesuaikan

$I_{\text{pengeluaran}}$  : indeks pengeluaran

$$IPM = \sqrt[3]{I_{\text{kesehatan}} \times I_{\text{pendidikan}} \times I_{\text{pengeluaran}}}$$

## B. Penelitian yang Relevan

Beberapa penelitian yang telah menggunakan analisis *cluster* K-Means dan Fuzzy C-Means antara lain.

1. Penelitian yang dilakukan oleh Ghosh dan Dubey (2013) dengan judul *Comparative Analysis of K-Means and Fuzzy C-Means Algorithms*. Data yang digunakan adalah *dataset Iris*. Hasil penelitian yang diperoleh adalah K-Means lebih baik daripada algoritma Fuzzy C-Means. Hasil *cluster* Fuzzy C-Means mendekati K-Means tetapi masih membutuhkan lebih banyak waktu komputasi daripada K-Means, hal ini dikarenakan keterlibatan perhitungan langkah-langkah *fuzzy* dalam algoritma.
2. Penelitian yang dilakukan oleh Panda *et al* (2012) dengan judul *Comparing Fuzzy C-Means and K-Means Clustering Techniques: A Comprehensive Study*. Dua ukuran jarak seperti *Manhattan* dan *Euclidean* digunakan untuk mencatat bagaimana ukuran jarak ini mempengaruhi kinerja pengelompokan secara keseluruhan. Kinerja telah dibandingkan berdasarkan tujuh parameter yaitu sensitivitas,

spesifisitas, presisi, akurasi, run time, jarak rata-rata intra *cluster* dan jarak antar *cluster*. Hasil penelitian diperoleh bahwa K-Means mengungguli Fuzzy C-Means berdasarkan 7 parameter tersebut.

3. Penelitian yang dilakukan oleh Sivarathri dan Govardhan (2014) dengan judul *Experiments on Hypothesis Fuzzy K-Means is Better Than K-Means for Clustering*. Penelitian dilakukan pada kumpulan data nyata yang diperoleh dari UCI *repository*. Kumpulan data yang digunakan terkait dengan penyakit diabetes. Hasil empiris mengungkapkan bahwa Fuzzy C-Means membutuhkan lebih banyak iterasi dan waktu jika dibandingkan dengan K-Means. Namun, akurasi dan kualitas *cluster* yang dibuat oleh Fuzzy C-Means lebih baik.
4. Penelitian yang dilakukan oleh Dubey *et al* (2018) dengan judul *Comparative Study of K-Means and Fuzzy C-means Algorithms on The Breast Cancer Data*. Hasil penelitian diperoleh bahwa algoritma Fuzzy C-Means memberikan hasil yang lebih baik dibandingkan dengan algoritma K-Means dikarenakan memiliki tingkat akurasi yang tinggi, dan memiliki hasil *cluster* yang lebih konsisten setelah dilakukan beberapa iterasi.

## BAB V

### PENUTUP

#### A. Kesimpulan

Berdasarkan hasil dan pembahasan yang diperoleh, dapat disimpulkan bahwa.

1. Pengelompokkan kabupaten/kota di Indonesia berdasarkan indikator penentu nilai IPM menggunakan metode K-Means dan Fuzzy C-Means dengan ukuran *cluster* yaitu 4 dengan capaian indikator IPM rendah, sedang, tinggi dan sangat tinggi. Capaian indikator IPM rendah terdiri dari 14 kabupaten/kota pada metode K-Means, pada metode Fuzzy C-Means terdiri dari 18 kabupaten/kota. Capaian indikator IPM sedang terdiri dari 175 kabupaten/kota pada metode K-Means, pada metode Fuzzy C-Means terdiri dari 183 kabupaten/kota. Capaian indikator IPM tinggi terdiri dari 237 kabupaten/kota pada metode K-Means, pada metode Fuzzy C-Means terdiri dari 222 kabupaten/kota. Capaian indikator IPM sangat tinggi terdiri dari 88 kabupaten/kota pada metode K-Means, pada metode Fuzzy C-Means terdiri dari 91 kabupaten/kota
2. Hasil *cluster* terbaik dalam mengelompokkan kabupaten/kota di Indonesia berdasarkan indikator IPM Tahun 2021 adalah metode Fuzzy C-Means.

## **B. Saran**

Berdasarkan hasil, pembahasan serta kesimpulan yang diperoleh, maka saran yang dapat diberikan adalah sebagai berikut.

1. Diharapkan peneliti selanjutnya menggunakan analisis *cluster* dalam pengelompokan berbagai permasalahan dengan menggunakan metode *cluster* yang lainnya.
2. Hasil penelitian yang sudah dilakukan dijadikan sebagai salah satu acuan dalam mengambil kebijakan supaya tidak terjadi keberagaman nilai IPM yang menyebabkan ketimpangan wilayah.

## DAFTAR PUSTAKA

- Alwi, W., & Hasrul, M. (2018). Analisis Kluster Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*, 6(1), 35-35.
- Badan Pusat Statistik (BPS). "Indikator Indeks Pembangunan Manusia Tahun 2021". Publikasi BPS diakses dari <http://www.sumbar.bps.go.id/>, tanggal 10 Januari 2022.
- Cebeci, Z., Yildiz, F., Kavlak, A. T., Cebeci, C., & Onder, H. Ppclust: *Probabilistic and Possibilistic Cluster Analysis*. 2019. URL: <https://CRAN.R-project.org/package=ppclust>, r package version 0.1, 3.
- Cebeci, Z., & Yildiz, F. (2015). Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures. *Journal of Agricultural Informatics*, 6(3), 13-23. <https://doi.org/10.17700/jai.2015.6.3.196>
- Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: an R package for determining the relevant number of cluster s in a data set. *Journal of statistical software*, 61, 1-36. <https://doi.org/10.18637/jss.v061.i06>
- Desgraupes, B. (2013). *Clustering indices*. University of Paris Ouest-Lab Modal'X, 1, 34.
- Dubey, A. K., Gupta, U., & Jain, S. (2018). Comparative study of K-means and fuzzy C-means algorithms on the breast cancer data. *International Journal on Advanced Science, Engineering and Information Technology*, 8(1), 18-29. <https://doi.org/10.18517/ijaseit.8.1.3490>
- Everitt, B. S. Landau, S., Leese, & M., Stahl, D. (2011). *Cluster analysis*. John Wiley & Sons.
- Ghosh, S., & Dubey, S. K. (2013). Comparative analysis of k-means and fuzzy c-means algorithms. *International Journal of Advanced Computer Science and Applications*, 4(4). <https://doi.org/10.14569/IJACSA.2013.040406>
- Hair, J., Black, W., Babin, B., & Anderson, R. (2013). *Multivariate Data Analysis. 7th ed.* Edinburgh: Pearson.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Hennig, C., & Imports, M. A. S. S. (2015). *Package 'fpc'*. Available at: Available at: <https://cran.r-project.org/web/packages/fpc/index.html> ENT, 91.